Claims

What is claimed is:

1. A method for accessing a memory device of a computer system, the memory device being electrically connected to a memory controller, the memory controller sequentially responding to a master device according to a sequence of access requests issued in order by the master device, the memory controller comprising a request queue, and a latency monitoring unit electrically connected to the request queue, the method comprising:

(a) using the request queue to store access requests generated from the master device;

(b) using the latency monitoring unit to record a plurality of latency values, the latency values respectively corresponding to the access requests stored in the request queue;

(c) using the memory controller to receive a first access request and add the first access request to the request queue with an associated queue priority according to latency values associated with the access requests already stored in the request queue; and

(d) using the memory controller to sequentially access the memory device according to the associated queue priorities of the access requests stored in the request queue.

2. The method of claim 1 wherein step (c) further comprises:

determining that the first access request is used to access a first page of the memory device; and

determining if a second access request used to access the first page of the memory device and a third access

request used to access a second page of the memory device have been stored in the request queue, the third access request immediately following the second access request and having a queue priority lower than a queue priority of the second access request.

3. The method of claim 2 wherein step (c) further comprises:

if a third latency value corresponding to the third access request is not greater than a maximum allowance value, assigning the first access request a queue priority that is higher than the queue priority of the third access request, assigning an initial value to a first latency value corresponding to the first access request, and increasing the third latency value by a predetermined increment value; and

if the third latency value corresponding to the third access request is greater than the maximum allowance value, assigning the first access request a queue priority that is lower than the queue priority of the third access request, and assigning the initial value to the first latency value corresponding to the first access request.

4. The method of claim 3 wherein the maximum allowance value is programmable.

5. The method of claim 3 wherein if the third latency value corresponding to the third access request is not greater than a maximum allowance value, the queue priority assigned to the first access request is lower than the queue priority of the

second access request.

6. The method of claim 2 wherein step (c) further comprises:

if the second access request stored in the request queue corresponds to a lowest queue priority, adding the first access request to the request queue with the lowest queue priority, and assigning an initial value to a first latency value corresponding to the first access request.

7. The method of claim 2 wherein step (c) further comprises:

if the request queue is empty, adding the first access request to the request queue, and assigning an initial value to a first latency value corresponding to the first access request.

8. The method of claim 1 further comprising:

respectively adding a predetermined increment value to the latency values corresponding to access requests in the request queue having associated queue priorities that are lower than the queue priority of the first access request.

9. The method of claim 1 wherein the memory device is a main memory of the computer system.

10. The method of claim 9 wherein the main memory is dynamic random access memory (DRAM).

11. The method of claim 1 wherein the memory controller is positioned within a north bridge circuit of the computer system.

12. A method of accessing a memory device of a computer system through a memory controller, the memory controller sequentially responding to a master device according to a sequence of access requests issued in order by the master device, the method comprising:

(a) storing access requests generated from the master device in a request queue;

(b) recording a plurality of latency values, the latency values respectively corresponding to the access requests stored in the request queue;

(c) receiving a first access request and adding the first access request to the request queue with an associated queue priority according to latency values associated with the access requests already stored in the request queue; and

(d) sequentially accessing the memory device according to the associated queue priorities of the access requests stored in the request queue.

13. The method of claim 12 wherein step (c) further comprises:

determining that the first access request is used to access a first page of the memory device; and

determining if a second access request used to access the first page of the memory device and a third access request used to access a second page of the memory device have been stored in the request queue, the third access request immediately following the second access request and having a queue priority lower than a queue priority of the second access request.

14. The method of claim 13 wherein step (c) further comprises:

if a third latency value corresponding to the third access request is not greater than a maximum allowance value, assigning the first access request a queue priority that is higher than the queue priority of the third access request, assigning an initial value to a first latency value corresponding to the first access request, and increasing the third latency value by a predetermined increment value; and

if the third latency value corresponding to the third access request is greater than the maximum allowance value, assigning the first access request a queue priority that is lower than the queue priority of the third access request, and assigning the initial value to the first latency value corresponding to the first access request.

15. The method of claim 14 wherein the maximum allowance value is programmable.

16. The method of claim 14 wherein if the third latency value corresponding to the third access request is not greater than a maximum allowance value, the queue priority assigned to the first access request is lower than the queue priority of the second access request.

17. The method of claim 13 wherein step (c) further comprises:

if the second access request stored in the request queue corresponds to a lowest queue priority, adding the first access request to the request queue with the lowest queue priority, and assigning an initial value to a first latency value corresponding to the

first access request.

18. The method of claim 13 wherein step (c) further comprises:

if the request queue is empty, adding the first access request to the request queue, and assigning an initial value to a first latency value corresponding to the first access request..

19. The method of claim 12 further comprising:

respectively adding a predetermined increment value to the latency values corresponding to access requests in the request queue having associated queue priorities that are lower than the queue priority of the first access request.

20. The method of claim 12 wherein the memory device is a main memory of the computer system.

21. The method of claim 20 wherein the main memory is dynamic random access memory (DRAM).

22. The method of claim 12 wherein the memory controller is positioned within a north bridge circuit of the computer system.

23. A memory controller for accessing a memory device of a computer system, the memory device being electrically connected to a memory controller, the memory controller sequentially responding to a master device according to a sequence of access requests issued in order by the master device, the memory controller comprising:

a request queue for storing access requests generated from

the master device;

a latency monitoring unit electrically connected to the request queue for recording a plurality of latency values, the latency values respectively corresponding to the access requests stored in the request queue; and

a reorder decision-making unit electrically connected to the request queue for controlling a first access request added to the request queue with an associated queue priority according to latency values associated with the access requests already stored in the request queue;

wherein the memory device is sequentially accessed according to the associated queue priorities of the access requests stored in the request queue.

24. The memory controller of claim 23 further comprises:

a page/bank comparing unit electrically connected to the reorder decision-making unit and the request queue for determining that the first access request is used to access a first page of the memory device and determining if a second access request used to access the first page of the memory device and a third access request used to access a second page of the memory device have been stored in the request queue, wherein the third access request immediately follows the second access request and has a queue priority lower than a queue priority of the second access request.

25. The memory controller of claim 24 further comprising:

a latency control unit electrically connected to the reorder decision-making unit and the latency monitoring unit for detecting whether a third latency value corresponding to the third access request is greater than a maximum

allowance value.

26. The memory controller of claim 25 wherein the maximum allowance value is programmable.

27. The memory controller of claim 23 wherein the memory device is a main memory of the computer system.

28. The memory controller of claim 27 wherein the main memory is dynamic random access memory (DRAM).

29. The memory controller of claim 23 being positioned within a north bridge circuit of the computer system.

30. A method for accessing a memory device of a computer system, the method comprising:

receiving one or more access requests for accessing the memory device in a first predetermined order; and

reordering the access requests in a second predetermined order to be processed in a request queue by relocating a first access request to follow a second access request accessing a same memory page,

wherein the relocating is prohibited if it increases a processing latency for a third access request to exceed a predetermined limit.

31. The method of claim 30 wherein the third access request immediately follows the second access request in the request queue before the first access request is inserted.

32. The method of claim 30 wherein the reordering further includes:

identifying that the first and the second access requests access the same memory page;

determining that the third access request is for accessing a different memory page; and

inserting the first access request between the second and the third access requests in the request queue.

33. The method of claim 32 further includes:

identifying a latency value associated with the third access request;

examining whether the latency value associated with the third access request is increased to exceed the predetermined limit if the first access request is to be inserted; and

increasing the latency value associated with the third access request by a predetermined incremental value if the first access request is inserted.

34. The method of claim 33 wherein the predetermined incremental value is programmable.

35. The method of claim 30 wherein the maximum latency value is programmable.

36. A method for accessing a memory device of a computer system, the method comprising:

receiving one or more access requests for accessing the memory device, the access requests accessing one or more memory pages; and

arranging the access requests in a predetermined order to be processed in a request queue by putting one or more access requests together for accessing a same memory page consecutively in one or more groups,

wherein an access request is prohibited from being grouped to be processed before at least one other access request if the grouping increases a processing latency of the at least one other access request in the request queue to exceed a predetermined limit.

37. The method of claim 36 wherein the arranging further includes:

identifying that a first access request received after a second access request accesses the same memory page as the second access request;

determining that a third access request is for accessing a different memory page;

inserting the first access request between the second and the third access requests in the request queue; and

repeating the above identifying, determining and inserting steps for all received access requests.

38. The method of claim 37 wherein the inserting further includes:

identifying a latency value associated with the third access request;

examining whether the latency value associated with the third access request is increased to exceed the predetermined limit if the first access request is to be inserted; and

increasing the latency value associated with the third access request by a predetermined incremental value if the first access request is inserted.

39. The method of claim 38 wherein the predetermined incremental value is programmable.

40. The method of claim 36 wherein the maximum latency value is programmable.